

Title	蛋白質折れ畳みダイナミックスにおける階層的規則性(複雑な多谷ポテンシャルエネルギー面上で生起する動力学的諸問題-力学的決定性と統計性の中間領域を探る(第2回)-,研究会報告)
Author(s)	松永, 康佑
Citation	物性研究 (2002), 78(4): 457-462
Issue Date	2002-07-20
URL	<a href="http://hdl.handle.net/2433/97252">http://hdl.handle.net/2433/97252</a>
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

## 蛋白質折れ畳みダイナミクスにおける階層的規則性

神戸大学理学部地球惑星科学科(非線形科学分野)

松永康佑 yasuo@kobe-u.ac.jp

蛋白質は如何にして効率的に native 状態を探し出すのか、というのは蛋白質の折れ畳みダイナミクスにおける重要な問題である。現在までのところ、蛋白質のポテンシャルエネルギー表面の地形に注目した研究が多く行われている。少数の研究者[1]を除いて、蛋白質の折れ畳みというのは、native 構造への bias があるエネルギーランドスケープ上で、“完全な”確率過程として生じ、いくつかのエネルギー障壁を乗り越えていくことで生じるものだと考えられてきた。本稿においては、我々はむしろ折れ畳みの動力的側面に注目する。蛋白質の熱的揺らぎの中に埋もれた規則性を、蛋白質の折れ畳みのシミュレーションから詳細に解析し、規則性と折れ畳みダイナミクスの効率性との関連性を議論する。

我々は、native 構造として  $\beta$ -barrel 構造をとる 46 個、3 種類の beads sequence で出来た off-lattice モデルを調べた[2]。ポテンシャルエネルギーは以下のように与えられる。

### 46 bead model

$$B_9N_3(PB)_4N_3B_9N_3(PB)_5P$$

B:hydrophobic, P:hydrophilic, N:neutral

$$E = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \{A(1 + \cos\Phi) + B(1 + \cos3\Phi)\}$$

$$+ \sum_{i < j-3} \left\{ 4\epsilon S_1 \left[ \left( \frac{\sigma}{R} \right)^{12} - S_2 \left( \frac{\sigma}{R} \right)^6 \right] \right\}$$

$$i, j, k \neq N: A = B = 1.2\epsilon$$

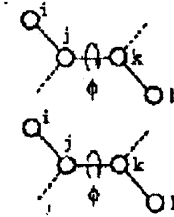
$$B \Leftrightarrow B: S_1 = S_2 = 1$$

$$P \Leftrightarrow B, P: S_1 = 2/3, S_2 = -1$$

$$N \Leftrightarrow B, P, N: S_1 = 1, S_2 = 0$$

$$\begin{cases} i, j = N \\ k, l \neq N \end{cases} \text{ or }$$

$$i, j, k = N: A = 0, B = 0.2\epsilon$$



このモデルに対して転移温度  $T_F$  とそれより高い・低い温度という、異なる 3 つの温度において等温分子動力学計算[3]を行った。解析の初めの段階では、それぞれの温度で主成分解析を行い、最小二乗的な意味で系の揺らぎを最も良く表している協調的な運動を抜き出した。これらの集団座標は、具体的には beads のデカルト座標の共分散行列を対角化して得られる。このようにして MD の結果として各主成分ごとにひとつの時系列が得られる。今回これらの時系列それぞれに対して“埋め込み論[4,5]”と呼ばれる非線形時系列解析を行った。埋め込み論とは、端的に言えば、一般的にある限定された時系列情報に対して、遅延座標系というものを使って、力学系の元来の状態空間と(位相的に)同じ性質をもった状態空間を再構成し、系の動力的特徴(次元性など)を抽出する理論・解析方法をいう。以下、例を挙げながら埋め込み論を説明していこう。

Lorenz アトラクターに対して埋め込み論を使ってみる。Lorenz アトラクターを記述する力学構造は既知であるが、ここでは、力学変数(もしくは  $X, Y, Z$  から成る任意の関数)のあるひとつ(例えば  $X(t)$ )しか観測する術がないとする。つまり、 $X(t)$  を観測された時系列情報として取り上げることとする(図 1)。オリジナルの状態空間の再構成をするために一般的に用いられている手法は、先ほど述べた遅延座標系

$$Y(i) = (X(i), X(i + \tau), X(i + 2\tau), \dots, X(i + (m - 1)\tau))$$

が使われる。ここで  $i$  は離散時間である。例として、3次元の遅延座標系の作り方を図2に示そう。このように、遅延座標系とは手に入れた時系列に対してある遅れ時間  $\tau$  を決めてやり、その整数倍の遅れ

時間の値を順に成分として持つ座標系である。こうして遅延座標系の次元を同様の方法で多次元に拡張していくと、ある次元 $m$ から、その遅延座標系で張られた空間は、時系列の背後にある系の状態空間と(位相的に)同じ性質をもった状態空間を再構成する。件のLorenzアトラクターの場合、3次元で似た構造が得られている(図3)。これが埋め込み論であって、この後につづく詳細な解析の舞台はこの再構成された多次元状態空間となる。埋め込み論に関してのいくつかの注意を述べておくと、まず、遅延時間 $\tau$ は数学的にはデータ数・精度が無限の条件の下では、“ほぼ”任意にとって良いことが許されているが、実データではこれらの条件は満たされることはなく、再構成の為に最適な $\tau$ を選ぶ必要が生じる。この $\tau$ の推定法としては数多くの方法が研究・提案されている。再構成に必要な最小の次元 $m$ (本稿ではこの $m$ を埋め込み次元と呼ぶことにする)の推定法についても種々の方法が提案されていて、統一的な方法はない。また $m$ の推定法は、実データの長さの有限性の為、超多次元の多様体の埋め込みについては必ず無力となる。こうした調べる系の自由度の大きさが実質的に制限されてしまう埋め込み論を、本稿のように蛋白質のモデルのような多体系に使う意義は何だろうか。最初に述べたが、本稿は蛋白質の熱的揺らぎの中に埋もれた規則性の発見・抽出をすることを目的としている。もし“ある重要な”自由度に関して、その運動が規則的であるならば、その時系列に対して埋め込みを行うことで、我々は縮小された新たな少数次元の空間の中で、状態空間を再構成し、そのダイナミクスの発展を予測する表現を得ることができるかもしれない。一方で、もし系の運動が“完全に”確率的なものであるならば、遅延座標系の性質上、常に再構成は不可能となる(実際は $m$ 推定法の限界として現れるだろう)。そこで本稿では、本来実用的見地から自由度の数が少ない場合にのみ有効であるはずの埋め込み論を、多体系の“ある”自由度(本稿では主成分を採用している)に注目して用いることで揺らぎの中に埋もれた規則性を発見・抽出する道具として使おうと考えている。

今回我々が行った埋め込み解析において、最も重要な点である埋め込み次元 $m$ 推定法について詳しく説明しよう。埋め込み次元の推定法としては、先ほど述べたように実に多くの方法が提案されている。この中から、今回我々が採用したのは、力学の解の一意性に矛盾する軌道の交差を調べていくという方法を用いたものである。これは、埋め込み論で証明されている、再構成された状態空間と元来の状態空間との間での一対一の関係を利用したものである。系の元来の運動が力学系であったならば、力学の解の一意性が存在している。そして、その性質は一対一の関係から再構成された状態空間においても成立しているはずである。つまり、ある次元で、力学系が埋め込まれているならばその再構成された空間において解の一意性は成り立っていないなければならない。そして、この解の一意性を使って埋め込み次元 $m$ を推定するには、Abarbanel[4]らによって提案されたfalse neighborsの概念が必要となる。ここで使っているfalseという言葉は“軌道の点同士が動力学的な理由で近くに存在するのではなく、遅延座標系の次元が低い(まだ埋め込まれていない)為に、高い次元から低い次元に投影されていて偶々近くにある”という意味において誤っている(false)と使われている。遅延座標系の次元が低い為に、力学の解の一意性が成り立っておらず、それゆえある点において交差が起こっているならば、その近傍に存在する点同士の関係はfalseであり、それらはfalse neighborsと呼ばれる。我々は、埋め込みを達成するまでの遅延座標系の各次元においてこのfalse neighborsの割合を見積もらなければならない。そして、false neighborsがゼロになった次元が埋め込み次元であると推定する。false neighborsの割合の見積もり方法は、具体的にはユークリッド距離などを使って計算する。ある次元において最近接にあった点同士がひとつ次元を上げたときの離れる距離が、他の最近接の点同士が離れる距離に比べて、大きければそれに対してfalse(これを最近接誤り点、false nearest neighbors FNNという)であるという判断をくだす。Abarbanelらの提案した方法は、このときある閾値を任意に決めて、それをfalseかtrueかの判断基準としなければならない。今回、我々が採用した方法はCao[6]による、こうした任意の閾値を使わずに推定する方法である。

Caoの方法について具体的にみていこう。実際にFNNを推定するためにCaoは以下の量を定義した。

$$a(i, d) = \frac{\|Y_{d+1}(i) - Y_{d+1}(n(i, d))\|}{\|Y_d(i) - Y_d(n(i, d))\|}$$

ここで、 $Y_d(i)$ は $d$ 次元の遅延座標系上の離散時間  $i$  でのベクトルであり、 $n(i, d)$ は $Y_d(i)$ に最近接の点の離散時間を表す。 $a(i, d)$ は、 $d$ 次元における $Y_d(i)$ の最近接点 $Y_d(n(i, d))$ が $d+1$ 次元になったときどれ程離れるかをみる量である。次元が上がるにつれて、全体として投影からくる交差をなくすことができたかどうかをみるには、次の $E1$ という指標が次元 $d$ 増加とともにどのように変わるかを見れば良い。 $E1$ は

$$E1(d) = \frac{E(d+1)}{E(d)}$$

と定義される。ここで、

$$E(d) = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} a(i, d)$$

である。最近接点同士の離れ方が次元 $d$ によらず均等になる(FNNがなくなる)とき、 $E1$ が1に収束する。つまり $E1$ が1に収束するときの次元を、我々は埋め込み次元と判定することができる。しかし、時系列の長さの有限性のために、例えば“全く”確率的につくりだされた時系列に対しても遅延座標系の次元が上がるにつれて、交差がなくなってしまうことがある。このとき、 $E1$ は力学系の時系列と同様に、ある高い次元で1に収束してしまう。このように時系列の長さが有限の場合、 $E1$ はある程度の高次元カオスと確率論的データとを区別することができない。こうした場合の為に、Caolはもうひとつの指標 $E2$ を導入した。 $E2$ は

$$E2(d) = \frac{E^*(d+1)}{E^*(d)}$$

と定義される。ここで、

$$E^*(d) = \frac{1}{N-d\tau} \sum_{i=1}^{N-d\tau} |X(i+d\tau) - X(n(i, d)+d\tau)|$$

である。 $E^*(d)$ は最近傍点が、 $d\tau$ 秒後に平均としてどれほど離れるかを計っている。もし $X(i)$ が“全く”確率的な時系列であったら、 $E^*(d)$ は次元 $d$ にほとんど依らずに定数を取り、 $E2$ の値は常に1付近であると期待できる。一方で、もし $X(i)$ の背後に力学系が存在するならば、小さい次元 $d$ の時には、まだFNNが多く存在する為に、 $E^*(d)$ は比較的小さな値をとるので、まだFNNが十分に排除されていない次元 $d$ では $E2$ の値は次元 $d$ に依るということが期待できる。FNNが十分に排除された後には、 $E2$ は次元 $d$ に依らなくなる。

埋め込み次元 $m$ 推定に関して、最後に先ほど述べた時系列の有限性による困難について詳しく説明しよう。ここでの説明を読んでいただくことで、読者の方々に埋め込み論を実際の未知の時系列に用いる場合における困難さが伝われば幸いである。先ほど述べた、そもそものfalse neighborsの概念には、前提として無限の長さの時系列が仮定されている。しかし、有限の長さ、数値上の時系列にはfalse neighborsで言われている近傍の概念は存在しない。そこで、先に述べたように我々は最近接のみの点を探してきて、その関係のみを推定することで誤った推定を避けている。しかし、もつと時系列の長さが足りない場合(このことさえ未知の時系列からは全く判断しかねるのだが)には、そもそもある次元で最近接を無理に探そうとすれば、それは時間的に近い為(時間相関があるため)に近いという(flow dataの時には)至極当たり前の点だけしか見つからなくなることもある。それを防ぐためにある時間間隔以内に入っている点であったならば探索しないという方法があるが、時系列によっては時間的相関と空間的相関という

時間スケールの区別をつけてその時間間隔を推定することは非常に難しく、見つかった最近接の点が果たしてfalse neighborsを見積もる上での候補に加えても良いのか判断に戸惑う次第となる。このような場合、flow dataに対してポアンカレ写像をとるという方法も考えられるが、データの量が劇的に減り解析は一層困難となるだろう。このように時系列の有限性からくる困難は、未知の時系列から運動の規則性を発見しようとする時には、そもそもの時系列の長さが本当に十分なのかどうか、上に述べたような問題を常に念頭におきながら解析をしなければ、間違った次元推定を導いてしまう。著者は以上の点が未知の時系列に対して、埋め込み解析を行う上での最も難しい、危険な点であると考えている。

今回の解析によって、系の揺らぎに大きく寄与している主成分から成る状態空間において、蛋白質が転移温度  $T_F(=0.72)$  付近では、埋め込み次元が他の 2 つの温度領域に比べて小さいことが分かった(極端に低い温度を除いて)(図 4)。この転移温度付近における運動の規則性は、Garcia らによる水溶液中のチクローム c の全原子 MD における解析結果を部分的に支持する。彼らによれば、水溶液中のチクローム c の揺らぎが大きい主成分上では、転移温度以下かつ 0.1ns 以上のスケールにおいて、平均二乗変位の時間依存ベキ指数  $\beta=1.75$  となる運動(super-diffusion)、すなわち力学との類推を用いれば、低次元におちる弾道的な運動が生じている。今回我々の埋め込み解析では、転移温度付近において、こうした弾道的な運動が捕らえられたのかもしれない。これらの結果を考察してみると、蛋白質の折れ畳みダイナミックスの概略図として図 5 のようなものを描くことができる。まず、高温では我々や Garcia らの結果( $\beta=1$  の通常拡散)が示唆するように、その運動は高い次元を伴う確率的ダイナミックスをしていると考えられる。また転移温度より低い温度領域の運動についても、埋め込み次元の高さから、依然そのダイナミックスの描像は高温領域とそれ程変わらず、小さい凸凹(rugged)なポテンシャルエネルギー曲面に捕獲され、そこでの抑圧・脱出のために  $\beta=0.5$  となる運動(sub-diffusion)や super-diffusion が生じているという拡散的描像で解釈される。それに対して転移温度付近では、埋め込み次元の低さが示唆するように、低次元な弾道的ダイナミックスが生じていて、その運動は通常の拡散よりも早く、その領域では蛋白質折れ畳みに対する従来の拡散的描像では決して捕らえきれない別の描像が存在する可能性が垣間見られる。

最近、Walesら[7]、Nymeyerら[8]によって、この46 bead modelはよく知られている"ファネル(funnel)型"エネルギー地形を必ずしも取っていないことが指摘されている。将来的には、native 構造における フラストレーション(frustration)がより小さいGō-likeモデルも併せて解析し[9]、エネルギーランドスケープの観点から、転移ダイナミックスの次元性・規則性などの動力学的相違と折れ畳みの効率の良さ・頑健性について議論する予定である。また他にも将来的には、媒質を陽に入れたときに、その媒質が蛋白質のダイナミックスの規則性に及ぼす役割の見極め、Garciaらの見つけた異なる層(0.1ns以上・以下)のダイナミックスを捕らえるための埋め込み論の展開、1分子計測など実際の時系列の解析、などを行いたいと考えている。

#### 【謝辞】

著者の指導教官である小松崎民樹助教授の日々の研究指導および励ましに感謝いたします。図 2、5 も氏に承諾の上借用させていただきました。併せて感謝いたします。多くの方々からの御批判、御意見をお待ちしております。

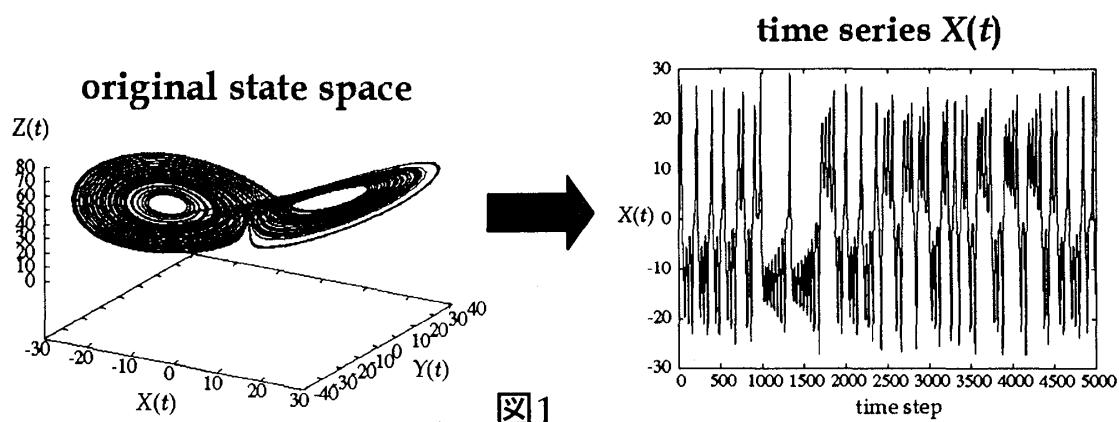


図1

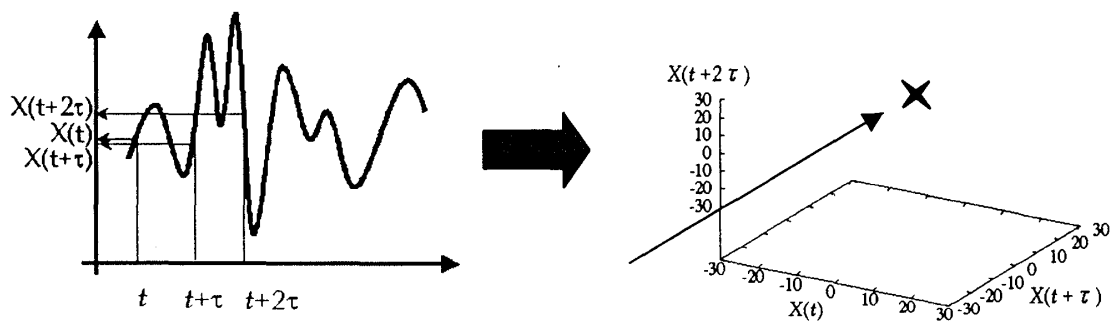


図2

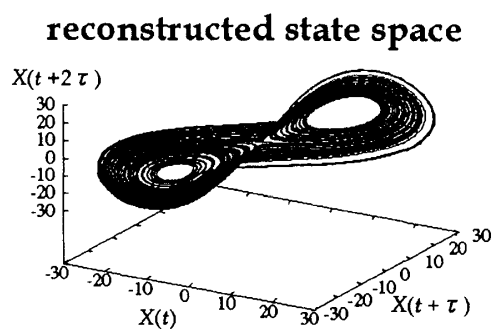


図3

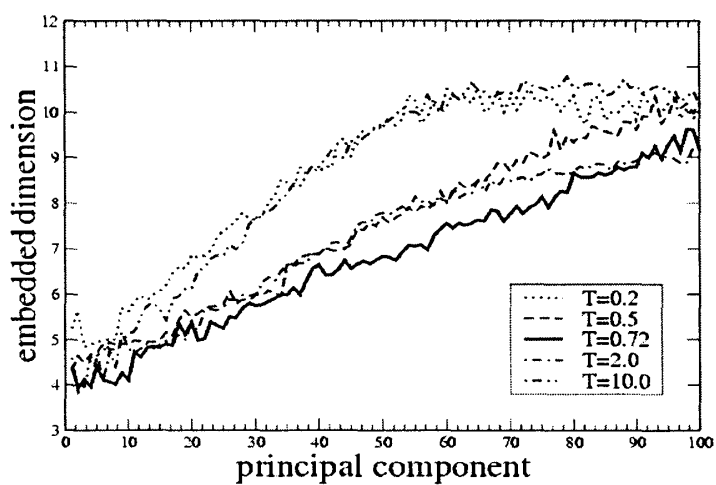


図4

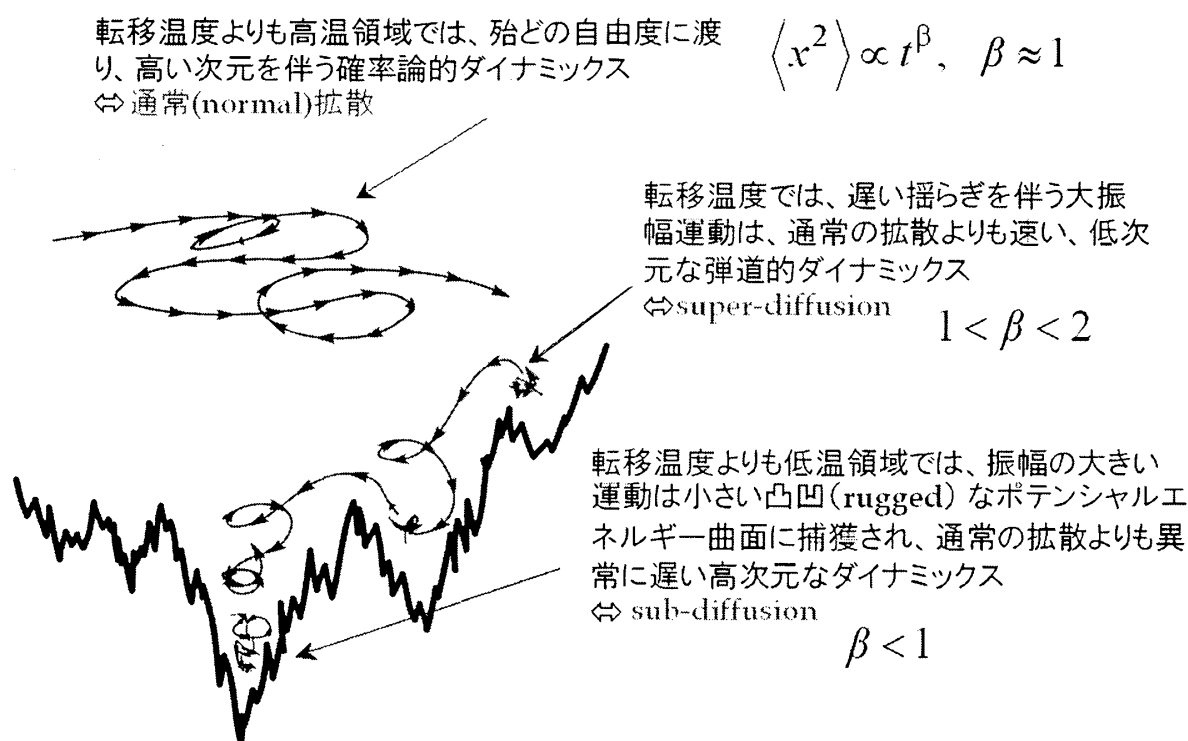


図5 タンパク質折れ畳みダイナミクスの階層的な規則性

【参考文献】

- [1] Steven S. Plotkin and Peter G. Wolynes *Phys. Rev. Lett.* **80** 5015 (1998).
- [2] J. D. Honeycutt and D. Thirumalai, *Biopolymers* **32**, 698 (1992).
- [3] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. *J. Chem. Phys.* **81**, 3684 (1984)
- [4] Henry D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer-Verlag New York, Inc. (1996).
- [5] Holger Kantz and Thomas Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press (1997).
- [6] Liangyue Cao, *Physica D*, **85** 225(1995).
- [7] David J. Wales, Jonathan P. K. Doye, Mark A. Miller, Paul N. Mortenson, and Tiffany R. Walsh *Adv. Chem. Phys.* **115**, 1 (2000)
- [8] H. Nymeyer, A. E. Garcia, and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **95**, 5921 (1998)
- [9] 小松崎民樹, 松永康佑 生物物理 投稿中 & Y. Matsunaga, K.S. Kostov, and T. Komatsuzaki, *J. Phys. Chem.* submitted for publication.